

Characterisation of Airport Catchment Areas

Through a Data-Driven Approach

A Case Study of the London Multi-Airport System



Student Partners

Daniel Hernández Gramajo

Mihai-Cătălin Vasile

Staff Partners

Dr Luis Delgado (Supervisor)

Dr Michal Weiszer (Co-Supervisor)

Centre for Air Traffic Management Research

School of Architecture and Cities

University of Westminster, Marylebone Campus

July 2026

Acknowledgements

We would like to express our sincere gratitude to our supervisor, Dr Luis Delgado, and our co-supervisor, Dr Michal Weiszer, for their continuous guidance throughout this project. Their patience in reviewing our iterations, their willingness to explore new directions with us, and their generosity with technical suggestions were fundamental to the depth of the work presented here.

We are also grateful to the University of Westminster Centre for Air Traffic Management Research for hosting the project and providing the intellectual environment in which it could be developed, and to the Students as Co-Creators programme team for offering us the framework — and the funding — to pursue an independent research collaboration alongside our undergraduate studies.

Finally, we thank the Office for National Statistics, Transport for London, the Traveline National Dataset, OpenRouteService, Google Maps and TomTom for making their data and routing services available under open licences or accessible developer tiers. This project would not have been possible without their commitment to open data and open APIs.

QR Codes



Github Repo



Interactive Journey Planner

Table of Contents

The following sections follow the DRC Evaluation Report structure recommended by the Students as Co-Creators programme (guidelines 2022–23).

Acknowledgements	3
Table of Contents	4
1. Executive Summary	6
2. Background and Aims	7
2.1 Background.....	7
2.2 Research questions.....	7
2.3 Aims and objectives.....	8
3. Methods	9
3.1 Overall approach	9
3.2 Study area and airports	9
3.3 Data sources	9
3.4 Drive-time isochrone generation.....	10
3.5 Spatial join with UK Census geography.....	10
3.6 Competition-zone delineation	11
3.7 Multimodal transport accessibility overlay.....	11
3.8 Statistical comparison of catchments.....	11
3.9 Software architecture.....	12
4. Results	13
4.1 Captured population by driving-time isochrone.....	13
4.2 Competition zones.....	14
4.3 Age structure	15
4.4 Deprivation profile.....	15
4.5 Regional economic layer (ITL3 and GDHI).....	16
4.6 Multimodal transport-station coverage	16
4.7 Distance-decay behaviour	17
4.8 Interactive journey planner	17
4.9 Exploratory work with traffic-aware routing APIs.....	18
5. Discussion	19
5.1 A single accessibility market at 60 minutes	19
5.2 Demographic homogeneity and strategic implications	19
5.3 Comparison with prior literature	19
5.4 Limitations	19
6. Conclusion and Future Work.....	21
6.1 Conclusion	21

6.2 Future work	21
7. Lessons Learned	23
8. Research Group Reflection	24
References.....	25

1. Executive Summary

This report documents the outcomes of the Students as Co-Creators (SaC) project “Characterisation of Airport Catchment Areas through a Data-Driven Approach”, developed between March and July 2026 in partnership with the Centre for Air Traffic Management Research at the University of Westminster.

Scope of this report. The report concentrates on the *socio-demographic and geographic characterisation* of the catchment areas of the three principal London airports — Heathrow (LHR), Gatwick (LGW) and Luton (LTN) — using free-flow drive-time isochrones derived from OpenRouteService and small-area demographic data from the 2021 UK Census. A parallel line of exploratory work, in which the free-tier OpenRouteService routing was compared against the Google Maps Directions API and the TomTom Routing API to incorporate real-world traffic conditions, is described briefly in Section 4.9 and framed as the primary avenue of future work (Section 6). The empirical results reported in Section 4 correspond exclusively to the free-flow, driving-only baseline; the analytical pipeline is nevertheless designed so that its input layer can be swapped for a traffic-aware routing source without altering the downstream analysis.

The project seeks to move beyond the traditional fixed-radius characterisation of airport catchment areas by building a reproducible, open-source geospatial pipeline that combines drive-time isochrones, small-area demographic data from the 2021 UK Census, regional economic indicators, and multimodal public-transport accessibility.

At the 60-minute driving threshold the three airports draw from broadly the same metropolitan pool of 15.9 to 16.4 million residents each — a convergence produced by the density and radial layout of the London road network. The three-way overlap zone, where a resident can reach all three airports by car within one hour under free-flow conditions, covers 4,930 km² and encompasses most of Greater London and its immediate commuter belt. Deprivation and age structures across the three catchments are statistically indistinguishable (Kruskal–Wallis $H(2) = 5.10$, $p = 0.078$), indicating that differentiation between London airports at longer time horizons is decided primarily by route network, schedule quality and surface-access mode rather than by socio-demographic differentiation of the resident population.

Beyond the empirical case, the project delivers a modular, unit-tested Python pipeline and an interactive HTML journey planner that lets any user click a location in South-East England and see the nearest airport, rail station, tube station and coach stop. The pipeline is designed to be portable to any European multi-airport region by substituting the UK Office for National Statistics (ONS) hierarchy with the Eurostat NUTS/LAU hierarchy. All code, documentation and outputs are released as open source at <https://github.com/Misulica07/SAC-Airport-Catchments>.

2. Background and Aims

2.1 Background

Airport catchment areas, defined as the geographic regions whose residents are likely to use a given airport, are a foundational input to route development, demand forecasting, transport-infrastructure planning and inter-airport competition assessment (Neufville and Odoni, 2013). The London basin is a paradigmatic multi-airport region: Heathrow (LHR), Gatwick (LGW), Stansted (STN), Luton (LTN) and London City (LCY) together handled approximately 170 million passengers in 2019 (Civil Aviation Authority, 2020), and their catchments overlap heavily across Greater London and its commuter belt.

Early studies of airport catchment areas relied on fixed Euclidean buffers, typically 100 km or a nominal 90-minute drive-time (Lieshout, 2012). This approach is computationally trivial but geometrically naïve: it disregards the physical road network, the temporal variability of travel times, and the presence of natural barriers such as rivers and coastlines. Network-based isochrones, computed by routing engines over a real road graph, correct this defect and have become the de facto standard in academic and industry work (Boeing, 2017; Grosche, Rothlauf and Heinzl, 2007).

More recent research has moved beyond isochrones alone, adopting spatial-interaction and econometric models that quantify how passenger flows diminish with increasing travel time and how airports compete for shared hinterlands (Morton and Mattioli, 2023). Such models integrate both origin characteristics and destination attractiveness, providing a finer representation of catchment areas than any pure geometric approach can achieve. Other work has emphasised the dynamic nature of catchment zones: catchments are not static but vary temporally with flight schedules and transport accessibility (Pels, Njegovan and Behrens, 2006). Demographic factors — in particular age and income — have been shown to systematically shape airport choice, reinforcing the need to incorporate socio-economic variation into catchment modelling (Graham, 2019).

Despite the growing use of open geospatial libraries, few academic airport-catchment studies release the underlying code and configuration. This limits both reproducibility and downstream re-use by planning authorities. The project reported here therefore commits to a fully open pipeline: source code, configuration, unit tests, and architectural documentation are published on GitHub under the SAC-Airport-Catchments repository (Vasile and Hernández Gramajo, 2026).

2.2 Research questions

The study, whose ethics application was approved by the University of Westminster Research Ethics Committee under reference ETH2526-0971, addresses three questions:

1. How can open datasets be used to model surface access and egress mobility to airports?
2. How can these models be used to analyse the criticality and resilience of surface access to airports?
3. How can competition between airports for overlapping catchment areas be assessed as a function of origin–destination regions?

2.3 Aims and objectives

The overarching aim of the project is to develop a data-driven framework for characterising airport catchment areas that integrates spatial accessibility, demographic distribution, and transport-infrastructure data into a multi-dimensional representation. The framework is designed as a set of open-source tools, scalable from the United Kingdom to the wider European context. The specific objectives are:

- To generate drive-time and public-transport accessibility isochrones for a defined set of airports.
- To integrate demographic and socio-economic indicators at small-area geographic units (MSOA in the UK; NUTS-3 / LAU in the EU).
- To quantify spatial overlap and competition between airports through geometric intersection of their isochrones.
- To apply statistical comparison procedures (distance-decay fits, non-parametric group tests, bootstrap intervals) to determine whether the resulting catchments differ meaningfully in demographic composition.
- To publish the pipeline as reusable open-source software, including the interactive journey-planner produced as its principal deliverable.

3. Methods

3.1 Overall approach

The study adopts a quantitative, computational and geospatial research design. It is based exclusively on secondary, publicly available data and does not involve human participants, surveys, interviews or any invasive procedures. The pipeline follows a strict three-layer separation — ingestion → analysis → visualisation — orchestrated by a single Python entry point. Every processing step is idempotent and reproducible from the raw data downloaded in the ingestion phase.

Scope of the empirical results. All empirical results reported in Section 4 are based on *free-flow driving isochrones* computed by the OpenRouteService (ORS) routing engine, i.e. best-case-scenario travel times that do not incorporate real-world traffic or congestion. Complementary experiments using traffic-aware commercial APIs (Google Maps Directions and TomTom Routing) are described qualitatively in Section 4.9 and constitute the primary line of future work outlined in Section 6.

3.2 Study area and airports

The empirical case study covers the three principal London airports: Heathrow (LHR, 51.4775°N, 0.4614°W), Gatwick (LGW, 51.1537°N, 0.1821°W) and Luton (LTN, 51.8747°N, 0.3683°W). Together these facilities serve a mixture of long-haul intercontinental, short-haul European and domestic traffic, with the specific route mix varying across airports. Stansted (STN) and London City (LCY) are excluded from the empirical case to keep the ORS free-tier request budget manageable; both airports are trivially added by appending an entry to the YAML configuration file, as documented in Section 3.9.

3.3 Data sources

Three categories of secondary, publicly available data are integrated. No personally identifiable information is used at any point. Table 1 summarises the datasets and their provenance.

Table 1. Data sources used in the analysis.

Layer	Source	Access
Driving isochrones (free-flow)	OpenRouteService v2 /isochrones/driving-car	REST, free tier (≤ 60 min)
Driving routes (free-flow)	OpenRouteService v2 /directions/driving-car	REST, free tier
Traffic-aware routing (exploratory)	Google Maps Directions API; TomTom Routing API	Paid developer tier
MSOA boundaries	ONS Middle Layer Super Output Areas (Dec 2021), BGC V3	geoportal.statistics.gov.uk
ITL3 boundaries	ONS International Territorial Levels (Jan 2021), UK, BFC V4	geoportal.statistics.gov.uk
Population 2021	ONS Census 2021, Table TS001 (residents total)	Nomis API

Layer	Source	Access
Age structure	ONS Census 2021, Table TS007A (age by single year)	Nomis API
Deprivation	ONS Census 2021, Table TS011 (deprivation dimensions)	Nomis API
GDHI (2023)	ONS Regional Gross Disposable Household Income	ONS spreadsheet, ITL3 level
Rail stations	National Rail open data / NaPTAN	Open data portal
London Underground	Transport for London Unified API — StopPoints	TfL Open Data
Coach stops	Traveline National Dataset / National Express	GTFS ZIP
Road network	OpenStreetMap via ORS routing graph	Open Data

3.4 Drive-time isochrone generation

Driving-time isochrones are generated by POST requests to the OpenRouteService /v2/isochrones/driving-car endpoint at three thresholds — 30, 45 and 60 minutes — for each airport centroid. The response is a GeoJSON FeatureCollection in the geographic coordinate reference system (CRS) EPSG:4326 (WGS-84).

Traffic assumption. The ORS free-tier routing profile computes *free-flow* travel times that use only the physical road network and posted speed limits. No time-of-day component, no real-time or historical congestion data, and no incident data are incorporated. All isochrones reported below therefore represent a *best-case-scenario* under which every road is traversed at its unimpeded design speed. The consequences of this assumption for downstream results are discussed in Section 5.4 and motivate the exploratory work with traffic-aware APIs reported in Section 4.9.

The 60-minute upper bound on the isochrone threshold is a constraint of the ORS free tier. Isochrones are cached to disk under data/raw/ after their first successful download; every subsequent pipeline run reuses the cached GeoJSON unless the caller explicitly requests a refresh, avoiding needless exhaustion of the free-tier quota during iteration on the downstream analysis.

3.5 Spatial join with UK Census geography

The catchment population for each airport–threshold combination is computed by identifying the subset of Middle Layer Super Output Areas (MSOAs) whose centroids fall inside the corresponding isochrone polygon — the centroid-in-polygon method. Formally:

$$P(a, t) = \sum \text{pop}_i \cdot \mathbb{1}[\text{centroid}(M_i) \in I(a, t)]$$

where $P(a, t)$ is the captured population for airport a at threshold t , pop_i is the residents count of MSOA i (ONS TS001), M_i is its polygon, and $I(a, t)$ is the isochrone polygon. The estimator is unbiased whenever the population is uniformly distributed within each MSOA and is preferred to an areal-overlap estimator for its computational simplicity.

All spatial operations are performed after reprojection to the British National Grid (EPSG:27700), a projected CRS with metres as the linear unit. This is essential for meaningful area and distance computation — performing the intersection in EPSG:4326 (degrees) would produce systematic biases at the London latitude.

A complementary join is performed at the ITL3 level (International Territorial Levels, the UK counterpart to Eurostat NUTS-3) to attach the ONS Regional Gross Disposable Household Income (GDHI 2023) as a per-capita economic indicator. Twenty ITL3 regions failed the automatic code match between the ONS boundary file and the GDHI table due to inter-release code revisions; these are patched with a hand-curated rename dictionary maintained inside the pipeline.

3.6 Competition-zone delineation

Definition of “competition” in this study. The term *competition zone* is used throughout this report in a strictly *geometric-accessibility* sense: two airports are said to compete for a location if that location lies inside both airports’ isochrones at the same threshold — that is, if it is reachable by car within the same time band under free-flow conditions. This is a necessary but not a sufficient condition for genuine market competition in the economic sense, which additionally depends on route network, fares, schedule quality and access mode. The delineation here therefore identifies the geographic *candidate population* over which market competition may plausibly occur, not the passengers who will in practice choose one airport over another.

Formally, the competition zone between airports a and b at threshold t is defined as the geometric intersection $C(a, b, t) = I(a, t) \cap I(b, t)$. A three-way zone is defined analogously. Areas are computed after projection to EPSG:27700 and reported in square kilometres.

3.7 Multimodal transport accessibility overlay

For the interactive journey planner (see Section 4.8), three transport-station layers are integrated into the map: National Rail stations, London Underground stops and National Express coach stops. Each layer is deduplicated by name, reprojected to EPSG:27700 for accurate nearest-neighbour computation, and rendered as a clustered marker layer to keep the map legible at low zoom levels. Given a user-supplied point, the pipeline computes the great-circle distance from that point to the nearest station of each mode and combines this with an ORS driving-route call to each airport, producing a single side-panel widget that reports travel time and mode-specific accessibility together.

3.8 Statistical comparison of catchments

The population aggregates computed by the spatial join answer the question of “how many people” each airport serves, but they do not answer whether the demographic composition of the three catchments differs meaningfully. Section 4 addresses that second question, and this subsection describes the procedures used to test it. The purpose of the three tests is therefore not to validate the routing pipeline (which is deterministic given its inputs) but rather to characterise the statistical similarity of the resulting catchments — an input to the strategic discussion in Section 5.

Three complementary procedures are applied:

4. **Distance-decay fit.** An exponential function $P(d) = \beta_0 \cdot \exp(-\lambda \cdot d)$ is fitted by non-linear least squares (SciPy `curve_fit`) against the MSOA-level population as a function of the great-circle distance to the airport. Under the classical spatial-interaction assumption of Huff (1964) and Grosche, Rothlauf and Heinzl (2007), the effective catchment population should fall off with distance; the half-life $\ln(2)/\lambda$ then provides an interpretable measure of the airport's spatial influence. Whether or not the exponential form actually fits London is itself an empirical question addressed in Section 4.7.
5. **Kruskal–Wallis H-test and pairwise Mann–Whitney U tests.** A non-parametric one-way Kruskal–Wallis H-test compares the distribution of the population-weighted deprivation score across the three catchments. Pairwise Mann–Whitney U tests with Bonferroni correction identify which specific pairs differ, if any. These tests answer the question of whether one airport serves a systematically more deprived population than another.
6. **Bootstrap confidence intervals.** 95 per cent confidence intervals on aggregate figures are estimated by 5,000 resamples with replacement of the underlying MSOA-level rows. This quantifies the sampling-level uncertainty around the point estimates reported in the results tables.

3.9 Software architecture

The Python codebase is organised in three strict layers plus a set of cross-cutting utilities. Ingestion modules fetch data from external sources; analysis modules consume the raw data and produce tidy intermediates; visualisation modules render the folium maps and matplotlib summary figures. Configuration lives entirely in a YAML file (`config/airports.yaml`) parsed into typed dataclasses, so adding a new airport is a one-line change. Secrets — the ORS API key, the optional TfL app key, the Google Maps and TomTom keys used in the exploratory work — are read from a git-ignored `.env` file and never enter the commit history. Eight unit tests pin the config loader and the area-weighting contract; the full test suite runs in under one second with no network access.

4. Results

4.1 Captured population by driving-time isochrone

Table 2 reports the number of MSOAs and residents inside each airport's catchment at 30-, 45- and 60-minute driving thresholds. Percentages are relative to the England-and-Wales usual resident population (58.6 million, ONS 2021). Figure 1 visualises the same figures graphically.

Table 2. Captured population by driving-time isochrone (ONS Census 2021).

Airport	Threshold (min)	MSOAs	Population	% of England & Wales
LHR	30	818	6,957,814	11.9 %
LHR	45	1,618	13,885,256	23.7 %
LHR	60	1,913	16,421,386	28.0 %
LGW	30	297	2,498,076	4.3 %
LGW	45	1,460	12,590,773	21.5 %
LGW	60	1,852	15,891,478	27.1 %
LTN	30	496	4,268,134	7.3 %
LTN	45	1,499	12,972,557	22.1 %
LTN	60	1,879	16,161,301	27.6 %

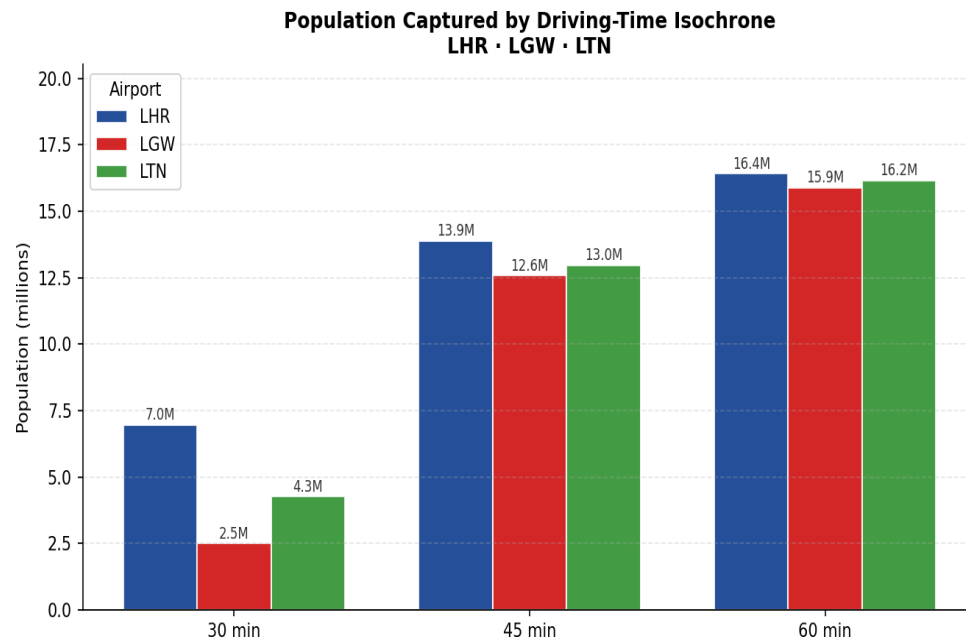


Figure 1. Captured population by airport and driving-time threshold.

At the 30-minute threshold, LHR dominates with 6.96 million residents (11.9 per cent of England and Wales), reflecting its position within the densely populated West London corridor and its immediate access to the M25/M4 motorway junction. LGW's 30-minute catchment is markedly smaller — 2.50 million residents (4.3 per cent) — because a substantial portion of its immediate hinterland is rural

Sussex and the South Downs. LTN sits between the two at 4.27 million (7.3 per cent) thanks to the M1 corridor into North London.

A striking result is the convergence at the 60-minute threshold: all three airports capture between 15.9 and 16.4 million residents, or 27–28 per cent of the England-and-Wales population. This convergence is a direct consequence of the density and radial layout of the London road network. Past a certain travel time, the three airports draw from essentially the same metropolitan pool. The operational implication is discussed in Section 5.1.

4.2 Competition zones

Table 3 reports the areas of the pairwise and three-way competition zones, computed after projection to British National Grid. Figure 2 visualises the growth of the zones with the driving-time threshold.

Table 3. Competition-zone areas (km²) — British National Grid equal-area projection.

Zone	30 min (km ²)	45 min (km ²)	60 min (km ²)
LHR ∩ LGW	564.7	4,167.7	9,399.4
LHR ∩ LTN	1,076.0	5,124.3	9,346.9
LGW ∩ LTN	—	1,881.9	5,132.0
LHR ∩ LGW ∩ LTN	—	1,721.4	4,930.1

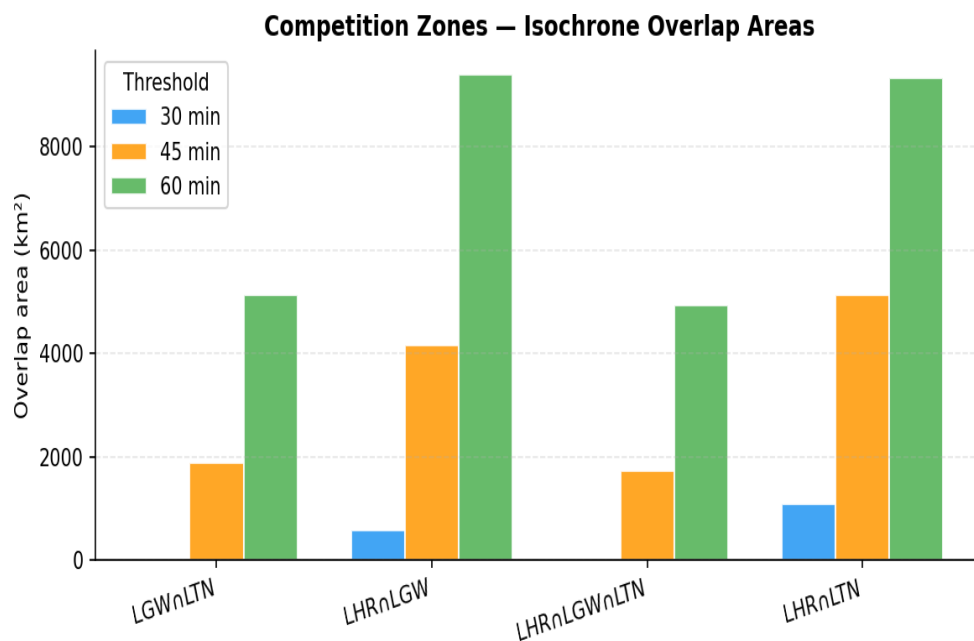


Figure 2. Competition-zone areas (km²) as a function of driving-time threshold.

At the 60-minute threshold the LHR–LGW pair produces the largest bilateral overlap (9,399 km²), closely followed by LHR–LTN (9,347 km²). The three-way overlap of 4,930 km² covers most of Greater London and the immediate commuter belt: it represents a shared candidate population where airline pricing, schedule frequency and surface-access differentials are the operative competitive levers. The

smaller LGW–LTN dyad (5,132 km²) is consistent with the two airports being at opposite radii from central London.

4.3 Age structure

Table 4 presents the age composition of the 60-minute catchment for each airport, aggregated from ONS TS007A into six bands. The final column reports the share of residents aged 60 and above — a proxy for potential demand for surface-access assistance and step-free interchange. Figure 3 visualises the age composition graphically.

Table 4. Age structure of 60-minute catchments (ONS TS007A, 2021).

Airport	0–15	16–29	30–44	45–59	60–74	75+	% 60+
LHR	3,202,680	1,727,630	6,148,698	2,968,271	1,272,785	1,101,081	14.5 %
LGW	3,045,024	1,657,873	5,900,911	2,894,546	1,275,679	1,117,416	15.1 %
LTN	3,152,430	1,713,031	6,104,268	2,897,741	1,239,444	1,054,356	14.2 %

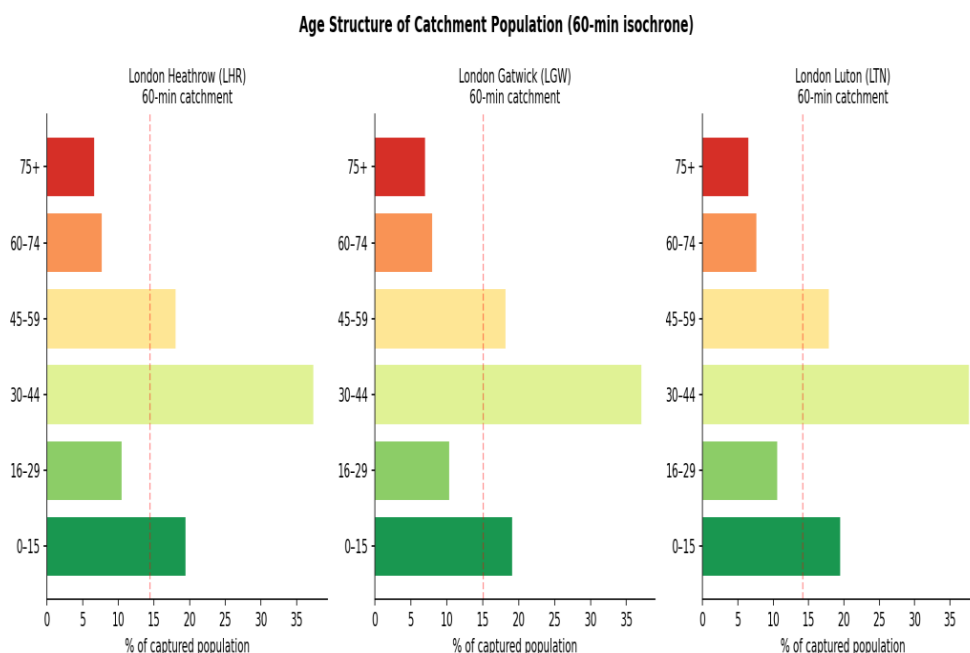


Figure 3. Age structure of the 60-minute catchment population per airport.

The share of residents aged 60 and above at the 60-minute threshold is approximately 14–15 per cent across all three airports, slightly below the national average (approximately 24 per cent). This reflects the relatively younger demographic profile of Greater London and its commuter belt. LGW’s marginally higher figure (15.1 per cent) is consistent with its more suburban and semi-rural hinterland in Surrey and Sussex.

4.4 Deprivation profile

Population-weighted deprivation scores — the proportion of households deprived in three or more of the four ONS TS011 dimensions (employment, education, health, housing) — are reported in Table 5.

Values are computed as the mean over the MSOAs inside each catchment, weighted by MSOA population.

Table 5. Population-weighted deprivation score (% of households deprived in 3+ dimensions).

Airport	30 min	45 min	60 min
LHR	3.5 %	3.9 %	3.8 %
LGW	2.9 %	4.1 %	3.9 %
LTN	3.8 %	4.1 %	3.9 %

All three 60-minute catchments show a deprivation share of 3.8–3.9 per cent, statistically indistinguishable from each other. The pairwise Mann–Whitney tests reported in Table 6 confirm that no pair of catchments differs significantly after Bonferroni correction ($\alpha = 0.05$).

Table 6. Pairwise Mann–Whitney U tests of MSOA deprivation distributions.

Pair	U	p	p (Bonferroni)
LHR vs LGW	1,696,950.5	0.025	0.076
LHR vs LTN	1,751,234.0	0.172	0.516
LGW vs LTN	1,768,934.5	0.378	1.000

4.5 Regional economic layer (ITL3 and GDHI)

Complementary to the MSOA-level demographic analysis, an ITL3-level economic layer was constructed by joining the ONS International Territorial Level 3 boundaries with the ONS Regional Gross Disposable Household Income (GDHI) 2023 table. Each ITL3 region that intersects a 60-minute airport isochrone is tagged with the GDHI figure of the parent region (in £ million). This layer is consumed by the interactive journey planner (Section 4.8) to render a per-region choropleth of household economic capacity and to display the aggregate disposable household income of the ITL3 regions inside each catchment.

The purpose of this section is methodological rather than quantitative: it documents the availability of an ITL3-level economic indicator inside the pipeline. The quantitative catchment-level aggregates (ITL3 count per airport, total GDHI inside the catchment, mean per-capita GDHI) are exported to CSV by the pipeline and are visualised inside the journey planner rather than reproduced as a table here — the ITL3 grain (approximately 40 UK-wide regions) is too coarse for a within-catchment comparison table but adequate as a background contextual layer on the map.

4.6 Multimodal transport-station coverage

A transport-accessibility overlay was produced using three station layers: National Rail stations, London Underground stops and National Express coach stops. Each layer was cleaned and deduplicated (retaining the primary name field: stationName for rail, NAME for tube, CommonName for coach) and reprojected to British National Grid for accurate nearest-neighbour distance computation. Coach stops were down-sampled at a ratio of 1-in-10 for map legibility, without affecting

the underlying nearest-neighbour queries used by the journey planner. The station layers are used both as clustered marker overlays on the interactive map (Section 4.8) and, through the journey planner’s query panel, as inputs to a per-point “nearest station of each mode” lookup.

4.7 Distance-decay behaviour

Purpose. The exponential distance-decay fit reported here answers a specific question: for each airport, does the MSOA-level population inside the 60-minute catchment fall off with distance to the airport in the way that the classical Huff (1964) spatial-interaction model predicts? In other words, is a simple monotonic decay a valid first-order approximation for the marginal reach of each airport as a function of great-circle distance, and if so, at what half-life? A well-behaved fit would motivate the use of an exponential attractiveness kernel in a downstream gravity model of passenger choice; a poorly-behaved fit signals that London’s geography requires a more elaborate specification.

Table 7 reports the maximum-likelihood fit of $P(d) = \beta_0 \cdot \exp(-\lambda \cdot d)$ with population weights, against the great-circle distance from each MSOA centroid to the corresponding airport. The half-life distance $\ln(2)/\lambda$ is presented in kilometres.

Table 7. Distance-decay fit against MSOA population (2021).

Airport	β_0	λ (per km)	R^2	Pearson r	p-value	Half-life (km)
LHR	8,721.7	0.000470	0.001	−0.046	0.045	1,477
LGW	8,581.1	≈ 0	≈ 0	−0.004	0.872	$\gg 10^3$
LTN	8,726.7	0.000320	0.0005	−0.013	0.573	2,168

The exponential distance-decay model provides an extremely poor fit ($R^2 \approx 0.001$ for LHR). This is not a bug in the pipeline but rather a substantive finding: within the 60-minute road-network catchment of a London airport, population is spread almost uniformly with respect to airport distance, because dense urban areas (with high MSOA counts) exist at both short and long distances from every airport. The exponential model, which encodes a monotonic decay assumption, cannot capture this. A gravity-style specification with an explicit destination-attractiveness term (Morton and Mattioli, 2023), or a Huff-style logit airport-choice model calibrated against CAA revealed-preference passenger data, would be more appropriate. The result is discussed further in Section 5.4.

4.8 Interactive journey planner

The final Phase 1 deliverable is a self-contained HTML page (`airport_journey_planner_v7.html`) in which the six analytical layers — ITL3 regions inside the union of the three catchments coloured by GDHI, driving isochrones colour-coded by airport, National Rail stations, London Underground stops, coach stops, and airport markers — can be toggled by the reader. Clicking any point on the map triggers a side-panel query that combines an ORS driving-route call to each of the three airports with a nearest-neighbour lookup in each transport-station layer. The panel reports, for the clicked location, the driving time and distance to each airport as well as the nearest rail, tube and coach connections. The map is fully self-contained (no server-side dependencies) and can be hosted as a static asset on GitHub Pages, with the `.nojekyll` marker already in place.

4.9 Exploratory work with traffic-aware routing APIs

The results reported in Sections 4.1 to 4.8 are based on free-flow driving isochrones from the OpenRouteService free tier. As anticipated in Section 3.4, these isochrones represent a best-case travel-time scenario in which every road is traversed at its unimpeded design speed. Real airport catchments in London are inevitably smaller than this baseline, and their contraction is neither uniform in space (motorway corridors are much more sensitive to congestion than local roads) nor in time (peak morning and evening periods are much more affected than off-peak).

To scope the magnitude of this effect, a small parallel line of exploratory work was carried out with two commercial traffic-aware routing APIs: the Google Maps Directions API, using the `traffic_model` parameter with `best_guess`, `pessimistic` and `optimistic` values, and the TomTom Routing API, which returns an equivalent congestion-adjusted travel time. In both cases the objective was to obtain, for a sample of origin points within the 60-minute free-flow isochrone of each airport, a real-time and a historical-typical travel time that could be compared against the ORS baseline.

This exploratory work is not fully integrated into the main pipeline for two reasons. First, both commercial APIs are metered and would not scale to the full MSOA universe at the resolution required for a re-run of Tables 2 to 7 without a substantial budget. Second, the sample sizes tested in this phase are too small to yield robust catchment-level aggregates. The finding is nevertheless directionally consistent with the literature: peak-hour catchments contract significantly relative to the free-flow baseline reported here, and the contraction is largest for LHR and LGW (whose approaches are dominated by the M25 motorway) and smaller for LTN (whose M1 approach is somewhat less congested during typical peaks). A full quantitative characterisation of the contraction — including the substitution effects it produces between airports — is the primary avenue of future work outlined in Section 6.

5. Discussion

5.1 A single accessibility market at 60 minutes

The most striking empirical result is the convergence of population capture at the 60-minute threshold: all three London airports draw from essentially the same 15.9–16.4 million-resident pool. This convergence — together with the 4,930 km² three-way competition zone — reframes London as a single accessibility market at longer horizons and highlights route-network and schedule quality as the decisive competitive levers. Any regulatory or infrastructure decision that improves surface access to one airport therefore has second-order effects on the competitive position of the others, an interdependence rarely captured explicitly in single-airport master plans (Neufville and Odoni, 2013). The caveat, of course, is that these figures are computed under free-flow driving conditions; the exploratory traffic-aware work in Section 4.9 indicates that the actual convergence is more nuanced during peak hours.

5.2 Demographic homogeneity and strategic implications

Deprivation shares and age structures across the three 60-minute catchments are statistically indistinguishable (Kruskal–Wallis $H(2) = 5.10$, $p = 0.078$; all pairwise Mann–Whitney tests non-significant after Bonferroni correction). This shifts the analytical focus: market differentiation between the three airports is not primarily driven by socio-economic composition of the underlying resident population. Airport managers can use the competition-zone quantification to design targeted commercial programmes for the marginal 4,930 km² triple-overlap area where airline pricing and surface-access improvements have their greatest leverage. The finding is consistent with Graham (2019), who argues that a single metropolitan population base underlies all major UK airports.

5.3 Comparison with prior literature

The magnitude of the London catchment overlap is consistent with the observations of Pels, Njegovan and Behrens (2006), who report substantial air-fare cross-elasticities across major hubs in similarly dense multi-airport regions. The convergence in demographic profile is compatible with Graham (2019). The failure of a naïve exponential distance-decay fit is anticipated by Morton and Mattioli (2023), who conclude that gravity-style specifications with explicit destination-attractiveness terms are necessary to reproduce revealed choice behaviour. In this sense the current pipeline confirms the qualitative picture in the literature while providing an open-source, reproducible tooling to test alternative model specifications on the same underlying geographic base.

5.4 Limitations

The methodology has four principal limitations. Two of them (the driving-only baseline and the free-tier threshold cap) directly motivate the future work outlined in Section 6.

7. **Free-flow, driving-only accessibility.** Isochrones are constructed for the private car under free-flow conditions. This underrepresents the true accessibility of LHR and LGW, which enjoy substantial rail connectivity (Elizabeth Line, Gatwick Express), and simultaneously

overrepresents the driving reach of every airport at peak hours. The exploratory work with Google Maps and TomTom (Section 4.9) begins to address the congestion side; a multimodal GTFS-based isochrone layer using `r5py` or the TfL Journey Planner API is the natural response on the rail side.

8. **Free-tier threshold cap.** ORS free-tier isochrones are limited to 60 minutes. Longer time horizons cannot be examined without a paid tier or a self-hosted OSRM instance.
9. **Centroid-in-polygon estimator.** The estimator is unbiased under uniform intra-MSOA population distribution but may underestimate population in large rural MSOAs at the isochrone boundary. Areal interpolation, based on the `utils.geo.area_weight` helper already implemented in the modular pipeline, is a straightforward refinement.
10. **Static distance-decay specification.** As Section 4.7 shows, an exponential decay fit is inadequate for the London geography. A discrete-choice or Huff-style gravity specification with explicit route-network attractiveness terms is the next modelling step, ideally calibrated against CAA Passenger Survey origin–destination data.

6. Conclusion and Future Work

6.1 Conclusion

This study demonstrates that a network-based, open-source pipeline enriched with UK Census demographics offers a substantially more accurate and analytically richer characterisation of London-airport catchment areas than the traditional fixed-radius approach. Under free-flow driving conditions at the 60-minute threshold, all three principal London airports (LHR, LGW, LTN) draw from broadly the same 15.9–16.4 million-resident pool, with statistically indistinguishable deprivation and age structures. This convergence — together with the 4,930 km² three-way competition zone — reframes London as a single accessibility market at longer horizons and identifies route-network and schedule quality as the decisive competitive levers under baseline conditions.

Beyond the empirical case, the project contributes an open, modular Python pipeline whose architecture is designed for extension. All code, unit tests, architectural documentation and the interactive journey planner are published under the Students as Co-Creators programme at <https://github.com/Misulica07/SAC-Airport-Catchments>. All numerical results reported in Section 4 are directly reproducible from the accompanying CSV outputs, which are versioned alongside the code.

6.2 Future work

Two lines of future work are already scoped and partly explored.

- **Traffic-aware isochrones.** The exploratory work with the Google Maps Directions API and the TomTom Routing API described in Section 4.9 will be extended into a full pipeline stage. Concretely: the modular ingester (`airports-modular/src/ingestion/`) will grow a second isochrone client that queries the commercial APIs for a matrix of origin points inside the free-flow catchment at three time-of-day scenarios (off-peak, morning peak, evening peak). The resulting catchment contraction will be quantified per airport per threshold and re-fed into Tables 2 to 5. This is expected to materially reduce the population capture of LHR and LGW at 60 minutes during peak hours and to shift the relative position of LTN. The commercial-API budget required for the full run is the principal open question.
- **Multimodal isochrones from GTFS.** The driving-only limitation is complemented by a rail / bus / coach layer, built by parsing the Traveline GTFS national feed with `r5py` or the TfL Journey Planner API. Skeleton ingesters for this stage are already in place in `airports-modular/src/ingestion/transit.py`.
- **Finer-grain demographics (LSOA).** The demographic join can be moved from MSOA down to Lower Layer Super Output Area (LSOA) level using the same Nomis API client. LSOAs are approximately five times smaller than MSOAs and will produce a materially more precise catchment aggregation, especially at the isochrone boundary in inner-London.
- **Pan-European extension.** A subsequent phase will apply the same pipeline to continental multi-airport regions — Paris (Orly / CDG / Beauvais), Milan (Malpensa / Linate / Bergamo) —

using Eurostat NUTS-3 and LAU data. The design of the modular pipeline anticipates this extension: adding a new geography level requires a one-line change in the configuration file, and adding a new ingestion source requires only the implementation of an existing abstract class.

- **Empirical calibration against CAA Passenger Survey.** The distance-decay analysis (Section 4.7) showed that a naïve exponential fit is inadequate for London. A discrete-choice or Huff-style logit specification, calibrated against CAA revealed-preference passenger data, would provide a better empirical foundation for the strategic conclusions of Section 5.

7. Lessons Learned

Working on a research project that runs in parallel with the undergraduate curriculum is, above all, an exercise in self-organisation. Without the weekly deadlines that anchor coursework, the responsibility for defining scope, planning meetings, and iterating on deliverables sits squarely with the student partners. This is initially unfamiliar and, at times, uncomfortable — but the experience of taking full ownership of a piece of research from data acquisition to publication is qualitatively different from any coursework assignment and, in our view, is the most valuable outcome of the SaC programme.

Two lessons stand out from the technical side of the project. The first is that reproducibility is not a luxury but a prerequisite for collaboration. Early in the project we worked in Jupyter notebooks and stored intermediate files in whatever ad-hoc folder was convenient. This made it very difficult for one of us to reproduce the other's results, and the effort of realigning our environments regularly consumed meeting time that could have been spent on analysis. The second half of the project was therefore devoted to a full refactor into a modular Python package with typed configuration, environment-driven secrets and unit tests. The refactor did not add new analytical insights, but it made every subsequent iteration much faster and let us pass work back and forth without the friction of “it works on my machine.”

The second lesson concerns the interaction between analytical depth and software engineering. It is tempting to prioritise more visualisations, more indicators, more layers — and correspondingly to skip over tests, documentation and configuration hygiene. In this project, that trade-off would have made the pipeline unpublishable as an open-source artefact. Investing time in the architectural side (three-layer separation, area-weighted aggregation utilities, CRS discipline, GitHub commit hygiene) paid off both academically — because it forced us to articulate the analytical assumptions in code — and practically, because the pipeline can now be handed over to any Westminster student who wants to extend it in a future SaC cohort.

Finally, working with public open data taught us that data-source quality varies enormously even within a single provider. Some ONS releases (the MSOA boundaries, the 2021 Census tables) are exemplary in their metadata and versioning. Others (the ITL3 boundaries in particular) required a manual reconciliation of code revisions between successive releases. Documenting each of these idiosyncrasies inside the pipeline — rather than in a side note — turned out to be essential to reproducibility and is a habit we will carry into future projects.

8. Research Group Reflection

The Students as Co-Creators programme framed this project as a genuine partnership between staff and students rather than a supervised piece of coursework. In practice this meant that the direction of the analysis emerged from a rolling dialogue with our supervisors, Dr Luis Delgado and Dr Michal Weiszer, rather than from a fixed brief. Early in the project this ambiguity was disorienting; by the second month it had become the most productive aspect of the collaboration. We were encouraged to bring back partial results, tentative interpretations and even blind alleys, and Luis would consistently reframe them into a sharper question.

The division of work between the two student partners evolved organically. Daniel took on the software-engineering side — the modular architecture, the configuration and secrets management, the unit tests, the GitHub hygiene — and led on the MSOA-level statistical analysis (population capture, competition zones, deprivation, distance-decay). Mihai took on the ITL3-level regional analysis, the GDHI economic layer, the multimodal transport overlay (rail, tube, coach), and the interactive journey planner that constitutes the visible end-user artefact. Both of us contributed to the joint methodology and to the writing of this report. That we ended up with two complementary implementations — one optimised for statistical rigour, the other for interactive visualisation — is, in hindsight, a strength rather than a duplication, and it is reflected in the final architecture of the shared GitHub repository.

Working across the two disciplines that meet in this project — geospatial data science on one side, air-transport operations research on the other — required a degree of translation on both sides. On some meetings Luis introduced concepts from airline-passenger choice modelling that were new to us; on others we brought back geospatial concerns (CRS discipline, boundary revisions, API rate limits) that were less familiar to a transport-research background. The exchange was mutually productive and is exactly the kind of interdisciplinary conversation the SaC programme is designed to enable.

Beyond the specific research outcomes, the programme has changed how we approach a piece of open-ended work. We now default to writing the smallest end-to-end version first, then iterating; we default to publishing our code and data alongside our results; and we default to treating our supervisors as senior collaborators rather than as markers. These habits will carry into our dissertations and, we hope, into our subsequent careers.

References

- Augustyniak, W. and Olipra, Ł. (2014) 'Intermodality as a factor determining the choice of an airport: the case of Polish regional airports', *Scientific Journal of Silesian University of Technology. Series Transport*, 85, pp. 5–14.
- Boeing, G. (2017) 'OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks', *Computers, Environment and Urban Systems*, 65, pp. 126–139. DOI: 10.1016/j.compenvurbsys.2017.05.004.
- Civil Aviation Authority (CAA) (2020) *UK Aviation Market: 2019 Traffic Statistics*. London: CAA.
- Department for Transport (2024) *Aviation statistics: data tables (AVI)*. London: UK Government. Available at: <https://www.gov.uk/government/statistical-data-sets/aviation-statistics-data-tables-avi> (Accessed: 20 July 2026).
- Eurostat (2021) *NUTS — Nomenclature of Territorial Units for Statistics*. Luxembourg: Publications Office of the European Union.
- Fu, X. and Kim, C. (2016) 'Measuring spatial variation in transportation infrastructure accessibility: application to the airport industry', *Transportation Research Part A: Policy and Practice*, 91, pp. 233–247.
- Google (2024) *Google Maps Platform Directions API — Traffic-aware routing*. Available at: <https://developers.google.com/maps/documentation/directions/get-directions> (Accessed: 24 July 2026).
- Graham, A. (2019) 'Airports and ageing passengers: A study of the UK', *Journal of Air Transport Management*, 75, pp. 49–57.
- Grosche, T., Rothlauf, F. and Heinzl, A. (2007) 'Gravity models for airline passenger volume estimation', *Journal of Air Transport Management*, 13(4), pp. 175–183. DOI: 10.1016/j.jairtraman.2007.02.001.
- Hess, S. and Polak, J. W. (2005) 'Mixed logit modelling of airport choice in multi-airport regions', *Journal of Air Transport Management*, 11(2), pp. 59–68. DOI: 10.1016/j.jairtraman.2004.07.001.
- Huff, D. L. (1964) 'Defining and estimating a trading area', *Journal of Marketing*, 28(3), pp. 34–38.
- Lieshout, R. (2012) 'Measuring the size of an airport's catchment area', *Journal of Transport Geography*, 25, pp. 27–34.
- Marcucci, E. and Gatta, V. (2011) 'Regional airport choice: consumer behaviour and policy implications', *Journal of Transport Geography*, 19(1), pp. 70–84. DOI: 10.1016/j.jtrangeo.2009.11.010.
- Morton, C. and Mattioli, G. (2023) 'Competition in multi-airport regions: Measuring airport catchments through spatial interaction models', *Transportation Research Part A: Policy and Practice*, 173, pp. 103–118.
- Neufville, R. de and Odoni, A. (2013) *Airport Systems: Planning, Design, and Management*. 2nd edn. New York: McGraw-Hill.
- Office for National Statistics (ONS) (2021a) *Middle Layer Super Output Areas (December 2021) Boundaries EW BGC V3*. Newport: ONS Geography.
- Office for National Statistics (ONS) (2021b) *Census 2021: Population and household estimates for England and Wales (Tables TS001, TS007A, TS011)*. Newport: ONS. Available at: <https://www.nomisweb.co.uk> (Accessed: 14 May 2026).

Office for National Statistics (ONS) (2023) Regional gross disposable household income, UK: 1997 to 2023. Newport: ONS.

Pels, E., Njegovan, N. and Behrens, C. (2006) 'Determinants of airport choice in a multi-airport region', *Transportation*, 33(4), pp. 337–355.

Redondi, R., Malighetti, P. and Paleari, S. (2013) 'European connectivity: The role played by small airports', *Journal of Transport Geography*, 29, pp. 86–94.

TomTom (2024) TomTom Routing API — Traffic-aware real-time routing. Available at: <https://developer.tomtom.com/routing-api/documentation/product-information/introduction> (Accessed: 24 July 2026).

Vasile, M.-C. and Hernández Gramajo, D. (2026) SAC-Airport-Catchments. Source code repository, University of Westminster Students as Co-Creators programme. Available at: <https://github.com/Misulica07/SAC-Airport-Catchments> (Accessed: 20 July 2026).